PriMera Scientific Engineering
Volume 7 Issue 3 September 2025
ISSN: 2834-2550

PriMera Scientific Publications

The Properties of the Least Squares Estimate in Regression Models with an Increasing Number of Unknowns Parameters

Type: Review Article
Received: August 01, 2025
Published: September 03, 2025

Citation:
Asaf Hajiyev., et al. "The Properties of the Least Squares Estimate in Regression Models with an Increasing Number of Unknowns Parameters". PriMera Scientific Engineering 7.3 (2025): 13-21.

Asaf Hajiyev[1]* and Jiuping Xu[2]

[1]*Institute of Control Systems, Ministry of Science and Education, Republic of Azerbaijan*

[2]*Business School, Sichuan University, Chengdu, China*

***Corresponding Author:** Asaf Hajiyev, Institute of Control Systems, Ministry of Science and Education, Republic of Azerbaijan.

Abstract

Regression models with an increasing number of unknown parameters and different and unknown values of variances of random errors of observations are considered. The specificity of such regression models is that there is no more than one response at each observation point, which does not allow estimation of the variances of random errors of observations. For the calculation of the l.s.e, the iteration process has been constructed. It is shown that the l.s.e are unbiased and consistent. Using these results, the approach for the construction of a confidence band for the unknown function in regression models is suggested.

Keywords: regression models with an increasing number of unknown parameters; least squares estimator; Gauss-Newton approach; iterative process

Introduction

In regression analysis, least squares is a parameter estimation method based on minimizing the sum of the squares of the residuals (a residual being the difference between an observed value and the fitted value provided by a model) made in the results of each equation. More simply, least squares is a mathematical procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets ("the residuals") of the points from the curve.

The most important application is in data fitting. When the problem has substantial uncertainties in the independent variable (the *x* variable), then simple regression and least-squares methods have problems; in such cases, the methodology required for fitting errors-in-variables models may be considered instead of that for least squares.

Least squares problems fall into two categories: linear or ordinary least squares and nonlinear least squares, depending on whether or not the model functions are linear in all unknowns. The linear least-squares problem arises in statistical regression analysis and has a closed-form solution. The nonlinear problem is usually solved by iterative refinement; at each iteration, the system is approximated by a linear one, and thus the core calculation is similar in both cases.

Regression analysis has a profound history, starting with observations over a short interval. The main idea is to get observations in short intervals, using scientific tools to make effective recommendations for practice. Such an approach was an impetus not only for solving practical problems but also contributed to the development of various fields of fundamental science.

The development and basis of regression analysis are related to the names of well-known scientists. The earliest form of regression is the method of the least squares, which was published by A. Legendre in 1805 and F. Gauss in 1809 in Theoria *Motvs Corporvm Coelestivm*. Two great scientists - A. Legendre and F. Gauss- applied the least squares method to determine, from astronomical observations, the orbits of bodies around the Sun (mostly comets, but also later newly discovered minor planets). Although A. Legendre published the first paper on the least squares in 1806, F. Gauss did it earlier. He created and used this method in 1795, when he was interested in astronomy, but did not publish it due to not paying serious attention to this problem.

However, the Academic world thinks that Friedrich Gauss was the creator of the least squares method and hence, a pioneer of regression analysis. Nevertheless, we think that the pioneers of regression theory are two distinguished scientists - Friedrich Gauss and Adrien-Marie Legendre. Friedrich Gauss [20, 22] and Adrien Marie Legendre [15] created the basis of regression analysis. Later, various scientists made essential contributions to the development of regression analysis, Fisher R., Huber P., Rao C.R., Seber G.F., and others [2, 13, 16, 19]. There are various types of regression models, including polynomials, robust, ridge, quantile regressions, and others [1, 3, 5, 10, 18]. In this paper, we consider linear and nonlinear models. The specificity of the models considered in this paper is that the variances of random errors of observations are different and unknown. Moreover, at each point of observation, there is no more than one response, which does not allow us to estimate the variances. Some authors try to estimate the variances of random errors of observations [7, 17, 23], but it is not possible in many practical tasks.

Regression models

Linear regression models

In statistics and various applications, linear regressions can be considered a statistical model that may be used for estimating the linear relationship between a scalar response and one or more explanatory variables. The case of one explanatory variable is called simple linear regression, and more than one, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable. If the explanatory variables are measured with error, then random errors in variables models are required, also known as measurement error models.

In linear regression analysis, the relationships are modeled using linear predictor functions whose unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, the conditional mean of the response given the values of the explanatory variables is assumed to be an affine function of those values; less commonly, the conditional median or some other quantile is used. Like all forms of regression analysis, linear regression focuses on the conditional probability distribution of the response given the values of the predictors, rather than on the joint probability distribution of all of these variables, which is the domain of multivariate analysis.

The most basic form of linear regression is known as simple linear regression, which is used to quantify the relationship between one predictor and one response variable. Assume that as a matter of experiment, there is observed some response y_i for which we have

$$y_i = f(x_i) + \varepsilon_i,\ i = 1,2,\ldots N. \qquad (2.1.1)$$

where $f(x)$ is some unknown function, ε_i is a random error of the observation, and N is the number of observations. If function $f(x)$ can be represented in the following form

$$f(x_i) = \theta_1\varphi_1(x_i) + \theta_2\varphi_2(x_i) + \ldots + \theta_m\varphi_m(x_i)$$

where $\theta_1, \theta_2,\ldots, \theta_m$ is the sequence of unknown parameters and $\varphi_1(x_i), \varphi_2(x_i),\ldots, \varphi_m(x_i)$ is the system of linearly independent functions,

$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ is the sequence of independent and identically distributed random variables (2.1.2)

The relation (2.1.1) is called a linear regression model with *m* unknown parameters. It is assumed that the number of unknown parameters *m* depends on the number of observations *N* and

$$\frac{m}{N} \to 0, \; N \to \infty \qquad (2.1.3)$$

The simple form of the linear regression model for 50 random points where the random errors have (normal) Gaussian distribution around the line $y = 1.5x+2$ is given in Fig.1. Thus, in Fig.1 is assumed that the data are approximated by the straight line $y = 1.5x+2$.

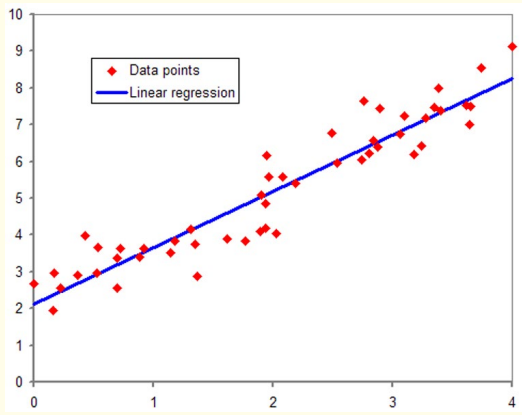

Figure 1: Linear regression model for 50 random points.

Nonlinear regression models

In statistics, nonlinear regression is a form of regression analysis in which observational data are modeled by a function that is a nonlinear combination of the model parameters and depends on one or more independent variables. The data are fitted by a method of successive approximations. Nonlinear models have quite complicated structures in comparison with linear models. In the capacity of examples of nonlinear functions can be considered the exponential functions, logarithmic functions, trigonometric functions, power functions, and others. Some functions, such as the exponential or logarithmic functions, can be transformed so that they become linear. When it is transformed into the standard linear regression can be performed but must be applied with caution. Unlike (2.1.1) in nonlinear regression, function $f(x)$ has a nonlinear structure that does not allow estimation of unknown parameters. For instance,

$$y_i = f(x_i, \theta) + \varepsilon_i, \; i = 1,2,\ldots N. \qquad (2.2.1)$$

where the function $f(x)$ nonlinearly depends on the parameter θ.

In Fig.2, there is an example of the nonlinear dependence of response from an independent variable. This data obeys a nonlinear law, i.e. nonlinear regression model can be used for the study of such a plot.

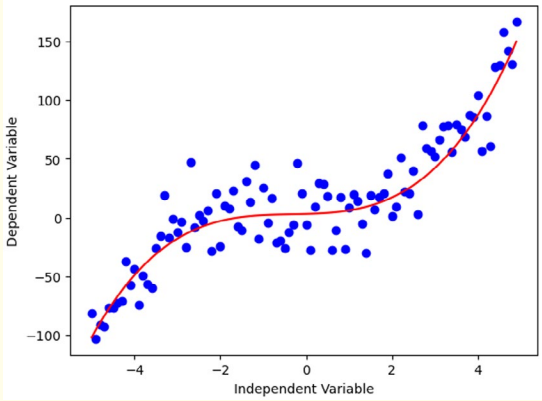

Figure 2: Nonlinear regression model.

The Least Squares Estimators

Least Squares Estimators for linear regression models

In practice, for a more precise estimation of the unknown function in regression models, it is necessary to increase the number of unknown parameters. Hence, regression models with an increasing number of unknowns are interesting from a theoretical as well as from a practical point of view. Such models are considered in [11, 12].

Consider a linear regression model (2.1.1). We can represent (2.1.1) in the following vector form $Y = X\theta$, where

$Y=\left(y_1, y_2, \ldots, y_N \right)$ is the vector of response,

$\theta=\left(\theta_1, \theta_2, \ldots, \theta_m \right)$ is the vector of the unknown parameters

$$X=\begin{pmatrix} \varphi_1(x_1) & \varphi_2(x_1) & \cdot & \varphi_m(x_1) \\ \varphi_1(x_2) & \varphi_2(x_2) & \cdot & \varphi_m(x_2) \\ \cdot & \cdot & \cdot & \cdot \\ \varphi_1(x_N) & \varphi_2(x_N) & \cdot & \varphi_m(x_N) \end{pmatrix}$$ is the design matrix

$\varepsilon=\left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N \right)$ is the vector of the unknown parameters and

$$E\varepsilon_i=0,\ E\varepsilon_i^2=\sigma_i^2 \le \sigma_0^2 < \infty$$

Where $\sigma_1^2, \sigma_2^2, .., \sigma_N^2$ are unknown and different (3.1.1)

At each point x_i, there is no more than one observation that does not allow us to estimate the variances of random errors.

(X^TX) is the Fisher information matrix.

For linear regression models, the l.s.e. based on N observations is defined as

$$\widehat{\theta}_N=(X^TX)^{-1}X^TY$$

Let us denote the eigenvalues of the matrix (X^TX/N) as

$0<\lambda_1(N) \le \lambda_2(N) \le \ldots \le \lambda_m(N)$, i.e., we assume that det (X^TX/N) ≠0

It is assumed that

$$\frac{m}{N\lambda_1(N)} \to 0,\ N \to \infty \qquad (3.1.2)$$

For vector a_N the expression $a_N \xrightarrow{P} 0$ (in probability) for $N \to \infty$ means that for any

$$\gamma > 0 \text{ the following ratio is held } P\left\{\left\|a_N\right\|^2 > \gamma\right\} \to 0 \text{ for } N \to \infty.$$

The main problem is defined when the l.s.e. converges on the true value of unknown parameters. Various authors investigated the issue of convergence of the l.s.e. to the true value of unknown parameters and suggested conditions when the l.s.e. is consistent. There is also the so-called BLUE (Best Linear Unbiased Estimators). It is shown in [8, 9] that the l.s.e. is an unbiased estimator with minimal variance among all linear estimators under some conditions. The consistency of the l.s.e. in linear regression models depends on the behavior of the eigenvalues of Fisher matrix.

For the linear regression model, the l.s.e. can be calculated directly through the Fisher matrix and vector of responses. Unfortunately for nonlinear regression models, the calculation of the l.s.e. requires some complicated steps. One of the effective approaches for the investigation of such models is the Gauss-Newton method [6], which allows the construction of the iteration process for the calculation of least square estimators. The main question is to find conditions when the iterated process, constructed by the Gauss-Newton method, converges to the true value of the unknown parameter. Introduce the following ratios:

a) $\sqrt{N}\left(\hat{\theta}_N - \theta\right) \xrightarrow{P} 0,\ (\textit{in probability})\ \textit{when}\ N \to \infty$

b) the components $\left(\hat{\theta}_i - \theta_i\right),\ i = 1, 2, \ldots, m(N)$; of the vector $\left(\hat{\theta}_N - \theta\right)$ have a normal distribution.

Theorem 3.1.1 Under the conditions of (2.1.1)-(2.1.3) and (3.1.2), the ratios a) and b) are held if and only if

$$\frac{tr\left(\frac{X^TX}{N}\right)}{N} \to 0,\ N \to \infty$$

Proof. If part. Let us consider

$$\hat{\theta}_N - \theta = \hat{\theta}_N - (X^TX)^{-1}X^T(X\theta + \varepsilon) = (X^TX)^{-1}\varepsilon.$$

Then it follows from here

$$\sqrt{N}\left(\hat{\theta}_N - \theta\right)^T\sqrt{N}\left(\hat{\theta}_N - \theta\right) = N\varepsilon^TX\left(\frac{X^TX}{N}\right)^{-1}\left(\frac{X^TX}{N}\right)^{-1}X^T\varepsilon$$

Using the Chebyshev inequality, we have

$$P\left\{/\left(\hat{\theta}_N - \theta\right)/ > \delta\right\} < \frac{E\left\{\varepsilon^TX\left(\frac{X^TX}{N}\right)^{-1}\left(\frac{X^TX}{N}\right)^{-1}X^T\varepsilon\right\}}{\delta} <$$

$$< \sigma_0^2 tr\left\{X\left(\frac{X^TX}{N}\right)^{-1}\left(\frac{X^TX}{N}\right)^{-1}X^T\right\} < \sigma_0^2 tr\left(\frac{X^TX}{N}\right)^{-1} \to 0,\ N \to \infty$$

according to the condition of Theorem 3.1.1.

Only if part. Assume that

$$P\left\{/\left(\widehat{\theta}_N-\theta\right)/>\delta\right\}\to 0,\ N\to\infty$$

Consider $\widehat{\theta}_N-\theta=\widehat{\theta}_N-\ (X^TX)^{-1}X^T(X\theta+\varepsilon)=(X^TX)^{-1}\varepsilon$

Then for the components of the vector, $(\widehat{\theta}-\theta)$ we have the following ratio

$$\left(\widehat{\theta}_i(N)-\theta_i\right)=\sum_{j=1}^{N}a_{ij}\varepsilon_j,$$

where $\widehat{\theta}_i(N)$ is *i-th* component of the vector $\widehat{\theta}$, a_{ij} are the elements of the matrix $(X^TX)^{-1}X^T$. For a_{ij}, according to [14] we have the following inequality

$$\max_{1\le j\le m}\sum_{i=1}^{m}|x_{i,j}|\le\frac{\sqrt{m}}{\lambda_1(N)}$$

As $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_N$ is the sequence of independent and identically distributed random variables, then according to the Central Limit Theorem [21], the $\left(\widehat{\theta}_i-\theta_i\right)$ components of the vector $(\widehat{\theta}-\theta)$ have a normal distribution.

Least Squares Estimates for Nonlinear Regression Models

Consider the nonlinear regression model (2.2.1). Unlike linear regression models in the nonlinear case, it is not possible to find the l.s.e. directly. For finding the l.s.e. for nonlinear regression models, it is necessary to construct an iterated process according to the Gauss-Newton approach [6].

Nonlinear regression models have more complicated structures because the unknown function has a nonlinear structure, which does not allow us to directly apply the methods used for the investigation of linear models.

Consider the following regression model

$$y_i=\eta(x_i,\theta^*)+\varepsilon_i\quad i=\overline{1,N}\qquad(3.2.1)$$

where $\eta(x_i,\theta*)$ - is nonlinear on θ function,

$\theta*$ - is the true value of the parameter θ and $\theta=(\theta_1,\theta_2,\ldots,\theta_N)^T$ is the vector of size N.

We assume that $\eta(x,\theta)$; $\frac{\partial\eta(x,\theta)}{\partial\theta}$; $\frac{\partial^2\eta(x,\theta)}{\partial\theta_i\partial\theta_j}, i,j=\overline{1,m}$ are continues on (x,θ) and bounded functions.

$\theta\in\Theta$ – is a compact set (3.2.2)

$E\varepsilon_i=0, E\varepsilon_i\varepsilon_j=r_{ij}, r_{ii}=\sigma_i^2, 0<\sigma_*^2\le\sigma_i^2\le\sigma_0^2$ (3.2.3)

Denote

$$f_{ij}(\theta)=\frac{\partial\eta(x_j,\theta)}{\partial\theta_i}, i=\overline{1,m}; j=\overline{1,N};$$

$\hat{\theta}$ – l.s.e.

$F_N(\theta)$-is the matrix of the experience with the elements $f_{ij}(\theta)$;

$0<\lambda_1^{(N)}(\theta)\le\lambda_2^{(N)}(\theta)\le\cdots\le\lambda_m^{(N)}(\theta)$ - are the ordered eigenvalues of the normed Fisher's matrix, $\left[\frac{F_N^T(\theta)F_N(\theta)}{N}\right]$,

$B(r)$-is the sphere with the radius $r>0$ with a center at the point $\theta*$.

For finding l.s.e. we can use the known iterated process, suggested in [5]

$$\theta(s+1) = \theta(s) + \left[F_N^T(\theta(s)) \cdot F_N(\theta(s))\right]^{-1} F_N^T(\theta(s)) \left(y - \eta(x, \theta(s))\right) \tag{3.2.4}$$

where all values of in (3.2.4) were defined above. The main problem is the convergence of the iterated process (3.2.4). Below, everywhere we assume that $\theta \in B(r)$.

$$\theta_N(s+1) = U(\theta_N(s)) = \theta_N(s) + A_N(\theta_N(s)) \cdot \delta_N(\theta_N(s)) \tag{3.2.5}$$

where

$$A_N(\theta_N(s)) = \left[\frac{F_N^T(\theta(s)) \cdot F_N(\theta(s))}{N}\right]^{-1} \frac{F_N^T(\theta(s))}{N} \tag{3.2.6}$$

$$\delta_N(\theta(s)) = y - \eta(x, \theta(s)); \;\; \delta^* = y - \eta(x, \theta^*) = \varepsilon$$

Denote

$$\zeta_{N,r}^n(\theta) = m \cdot \frac{\partial A_N(\theta)}{\partial \theta_n} \cdot \varepsilon, \theta \in B(r), n = 1,2, \dots, m.$$

Theorem 3.2.1 Assume that the conditions (3.2.2) and (3.2.3) are held.

If there exists such N that the following conditions (3.2.10) , (3.2.11) are held

$$\frac{m^4\sqrt{m}}{N(\lambda_1(\theta))^3} \to 0 \;\; \text{for} \;\; r \to 0 \tag{3.2.10}$$

$$\frac{m^5}{N(\lambda_1(\theta))^4} \to 0 \;\; \text{for} \;\; r \to 0 \tag{3.2.11}$$

$$\text{Then } \zeta_{N,r}^{(n)} \xrightarrow{P} 0 \quad \text{for } r \to 0$$

Proof. From the matrix analysis [11] for $\frac{\partial A_N}{\partial \theta_k}$, $k = 1, 2, .., m$; we have the following ratio

$$\frac{\partial A_N(\theta)}{\partial \theta_k}\varepsilon = \frac{1}{N}\left[\frac{F_N^T(\theta(s))F_N(\theta(s))}{N}\right]^{-1} \frac{\partial F_N^T(\theta)}{\partial \theta_k} + \frac{1}{N}\left[\frac{F_N^T(\theta)F_N(\theta)}{N}\right]^{-2} \times$$

$$\times \left[\left(\frac{\partial F_N^T(\theta)}{\partial \theta_k} F_N(\theta) + F_N^T(\theta)\frac{\partial F_N(\theta)}{\partial \theta_k}\right)/N\right] \cdot \varepsilon \tag{3.2.13}$$

Denote $I_k, i = 1, 2, 3$; k-th term in (3.2.13). Then

$$I_1 \le m \cdot \frac{C}{N^2} \cdot m\left(\frac{\sqrt{m}}{\lambda_1(\theta)}\right)^2 \cdot \left(N + \sum_{i \ne j} |r_{ij}|\right) =$$

$$= C_1\left(\frac{m^3}{N(\lambda_1(\theta))^2} + \frac{m^3}{N^2(\lambda_1(\theta))^2}\sum_{i \ne j}|r_{ij}|\right) \to 0, \qquad for \;\; r \to 0$$

according to the condition (3.2.11).

Opening parenthesis in I_2 and denoting $J_{2,i}(\theta)$ i-th term in I_2 we have

$$J_{2,1}(\theta) = \frac{m}{N^2} E\left[\varepsilon^T \frac{\partial F_N(\theta)}{\partial \theta_k}\left(\frac{F_N^T(\theta)F_N(\theta)}{N}\right)^{-3} \frac{\partial F_N(\theta)}{\partial \theta_k} F_N(\theta) F_N^T(\theta)\varepsilon\right]$$

Similarly, we have

$$J_{2,1}(\theta) \le \frac{C}{N^2} \cdot m^3 \left(\frac{\sqrt{m}}{\lambda_1(\theta)}\right)^3 \left(N + \sum_{i \ne j} |r_{ij}|\right) =$$

$$= C \cdot \frac{m^4\sqrt{m}}{N\left(\lambda_1(\theta)\right)^3} + \frac{m^4\sqrt{m}}{\left(\lambda_1(\theta)\right)^3} \cdot \frac{1}{N^2} \sum_{i \ne j} |r_{ij}| \to 0, r \to 0,$$

under the condition (3.2.11).

By the same approach, we can prove $J_{2,2}(\theta) \to 0$ and $I_3 \to 0$ for $r \to 0$.

Open parenthesis in I_4. Denote $J_{4,i}$ i-th term in I_4.

$$J_{4,1} = \frac{m}{N^2} E\left(\varepsilon^T F_N(\theta) \frac{\partial F_N^T(\theta)}{\partial \theta_k} F_N(\theta) \left(\frac{F_N^T(\theta) F_N(\theta)}{N}\right)^{-4} \frac{\partial F_N^T(\theta)}{\partial \theta_k} F_N(\theta) F_N^T(\theta) \varepsilon\right)$$

According to Lemma 3.2.1 and the conditions of Theorem 3.2.1 we have

$$J_{4,1} = \frac{1}{N^2} \cdot C \cdot m^3 \left(\frac{\sqrt{m}}{\lambda_1(\theta)}\right)^4 \left(N + \sum_{i \ne j} |r_{ij}|\right) =$$

$$= \frac{C}{N} \cdot \frac{m^5}{\lambda_1^4(\theta)} + \frac{m^5}{N^2 \lambda_1^4(\theta)} \sum_{i \ne j} |r_{ij}| \to 0, r \to 0.$$

Other terms of $J_{4,i}$; i = 2, 3, 4; have the same form as $J_{4,i}$. Hence, $J_4(\theta) \to 0$ for $r \to 0$. Then from Chebyshev's inequality we have

$$P\left\{\left\| m \cdot \frac{\partial A_N(\theta)}{\partial \theta_k} \varepsilon \right\|^2 > a\right\} \le \frac{E\left(\varepsilon^T \frac{\partial A_N^T(\theta)}{\partial \theta_k} \cdot \frac{\partial A_N(\theta)}{\partial \theta_k} \cdot \varepsilon\right)}{a^2} \le$$

$$\le \frac{I_1 + I_2 + I_3 + I_4}{a^2} \to 0, r \to 0$$

which proves Theorem 3.2.1

Conclusion

In the paper, the method for calculating the l.s.e. is suggested. Under some conditions for eigenvalues of the Fisher matrix, it is proved that the l.s.e. is unbiased and consistent. For nonlinear regression models, the Gauss-Newton is constructed. It is shown that if the first approximation in the Gauss-Newton method is taken from the sphere of radius r>0, then for $r \to 0$, $s \to \infty$, (s is a number of the iterations), the solution of the Gauss-Newton iterative process converges to some point θ_N and in the capacity of the l.s.e. can be taken

$$lim_{N \to \infty} \theta_N = \widehat{\theta}$$

Such estimators can be used in the capacity of unknown parameters and for estimating the elements of a covariance matrix of the deviation vector. This process can allow us to construct a confidence band for an unknown function in a regression model. The details investigation of such regression models can be found in the book [24], which will be published in the end of 2025.

References

1. MK Acharjee and K Das. "Polynomial Spline Regression: Theory and Application". Statistics > Methodology (2023).

2. J Aldrich. "Fisher and Regression". Statist. Sci 20.4 (2005): 401-417.
3. A Belloni and V Chernozhukov. "1-penalized quantile regression in high- dimensional sparse models". The Annals of Statistics 39 (2011): 82-130.
4. RA Bellman. "Introduction to matrix analysis". NY, McGraw-Hill (1962).
5. PL Davies. "Aspects of robust linear regression". Ann. Statist 21 (1993): 1843-1899.
6. E Demidenko. "Mixed Models". Theory and Applications with R. Wiley Series in Probability and Statistics (1981).
7. LY Deng and CFJ Wu. "Estimation of variance of regression estimator". J.American Statist. Assoc 82 (1987): 568-576.
8. H Drygas. "BLUE in linear regression models". Probab.Mathem. Statist.. Fasc 2 (1985): 177-186.
9. H Drygas. "Weak and strong consistency of least squares estimators in linear models". Z. Probability theory. Administrative (1976): 119- 128.
10. L Fahrmeir., et al. "Regression". Models, Methods and Applications. Springer (2021).
11. AH Hajiyev. "Regression Models with Increasing Numbers of Unknown Parameters". In: International Encyclopedia. Statist. Sciences, Springer (2011): 1208-1211.
12. AH Hajiyev and J Xu. "Gauss-Newton method for a nonlinear regression model with an increasing number of unknown parameters". Azerbaijan J. Mathematics 15 (2025): 2.
13. PJ Huber. "Data analysis. What can be learned from the past 50 years". Wiley. Probab.Statist (2011).
14. VI Krylov, VV Bobkov and PI Monastirski. "Theory of computing methods". Minsk, Nauka I Tekhnika (1985). (in Russian).
15. A-M Legendre. "On Least Squares". Translated from the French by Professor Henry A Ruger and Professor Helen M Walker, Teachers College, Columbia University, New York City (1805).
16. CR Rao. "Linear Statistical Inference and its Applications". John Wiley Series in Probability and Statistics.
17. Rao C.R. "Estimation of heteroscedastic variances in linear regression". J. American Statistician. Association 65 (1970): 161-172.
18. PJ Rousseeuw and A Leroy. "Robust regression and outlier detection". Wiley (1987).
19. GF Seber. "Linear Regression Analysis". NY, Wiley Ser. Probab. and Statistics (2003).
20. O Sheinin. "Gauss and the method of least squares". Yearbooks of Economics and Statistics (1999).
21. AN Shiryaev. "Probability". Springer-Verlag, New York-Berlin--Heidelberg (2013).
22. S Stigler. "Gauss and the Invention of Least Squares". Ann. Statist 9.3 (1981): 465-474.
23. Zhang Xiaoqin, Hao Hongxia and Liang Jiye. "A new nonparametric estimation method of the variance in a heteroscedastic model". Hacettepe Journal of Mathematics and Statistics 44.1 (2015): 239-245.
24. A Hajiyev and J Xu. "High-dimensional regression analysis and Artificial Intelligence". Springer (2025).